Capítulo
1

A Bioestatística e a Preparação de Estudos

1 Conceitos Preliminares

1.1 Estatística

1.1.1 Conceito e objetivo

Entende-se por estatística a área do conhecimento que se encarrega especificamente da coleção ou da reunião de dados.

O objetivo de reunir dados é o de fornecer ***informações*** sobre as características de grupos de pessoas ou coisas (Fig. 1.1). As informações têm por objetivo conhecer algum aspecto (ou questão) relacionado a esses grupos e, desta forma, servir de base para a escolha dos procedimentos mais adequados para resolvê-lo.

Por exemplo, quando um cardiologista solicita do seu paciente informações referentes ao seu histórico familiar sobre doenças cardiovasculares, DCV, está levantando um dado que já mostrou, a partir de dados referentes a outros pacientes, apresentar uma possível relação com o seu prognóstico cardiológico. Essa e outras informações, como tipo de alimentação, exposição a tensões e sedentarismo, irão auxiliar a compor um quadro dos fatores que podem contribuir para melhorar ou prejudicar a saúde do paciente. Essas informações são de natureza estatística, aplicadas, neste caso, à medicina.

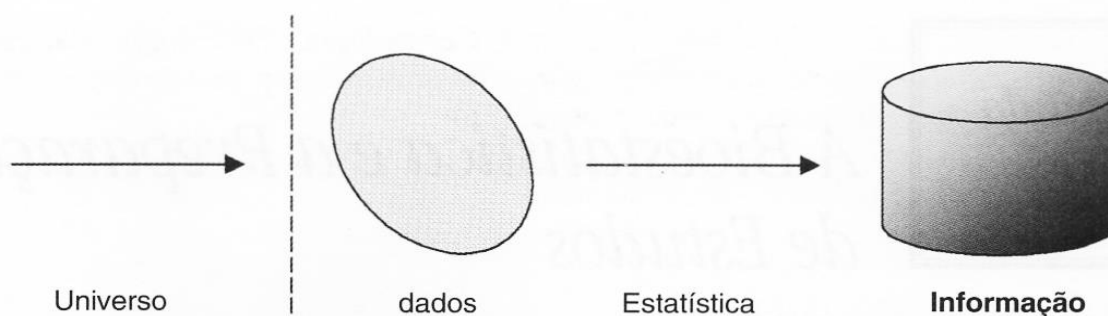



Fig. 1.1 Metodologia da estatística descritiva.

Evidentemente, trata-se de fornecer a informação da forma mais inteligível e completa possível. Desta forma, são utilizadas rotinas e meios que permitam um bom entendimento das informações, ***organizando*** os dados. A organização de vários grupos de dados dá origem aos ***bancos de dados***.

Define-se como primeiro (porém não mais importante) objetivo da estatística tornar a informação clara e precisa ao receptor, valendo-se do ferramental disponível. Atualmente, os recursos automáticos e gráficos da microcomputação são ferramenta indispensável para o tratamento da informação e, por extensão, para a estatística.

1 1.2 Divisão

Costuma-se dividir a estatística em duas partes:

- **Descritiva:** encarrega-se do levantamento, organização, classificação e descrição dos dados em tabelas, gráficos ou outros recursos visuais, além do cálculo de parâmetros representativos desses dados.
- **Analítica:** trabalha com os dados de forma a estabelecer hipóteses em função desses dados, procede a sua comprovação e, posteriormente, elabora conclusões científicas.

Uma terceira parte, que será aqui denominada ***Estatística de Planejamento***, poderia ser acrescentada e refere-se, basicamente, à otimização estatística das ações postuladas em função das conclusões obtidas na fase analítica. Esta parte é denominada *Teoria da Decisão Estatística* ou *Estatística Bayesiana.*

1.2 Bioestatística

Bioestatística é a estatística aplicada às ciências que estudam aspectos vitais (referentes à vida), como Medicina, Biologia, Nutrição, Fisioterapia, Odontologia, Farmácia, Psicologia, Enfermagem, Veterinária, Agronomia, Engenharia Ambiental e outras.

Na Medicina, especificamente, pode ser entendida em dois ambientes. O primeiro, referente ao levantamento de informações, como registro de doenças, surtos, endemias, epidemias, e de registros de qualidade de vida, como condições de alimentação, sanitárias, habitacionais, de prevenção de doenças, educação etc. Denomina-se esse ambiente de ***ambiente macro***, e tem a ver fundamentalmente com a identificação, a planificação e a execução de ações de ***Saúde Pública***. Neste caso, constitui-se num ferramental fundamental para cadeiras do curso de Medicina como ***Epidemiologia, Medicina Preventiva, Organização de Sistemas de Saúde*** etc.

O segundo ambiente refere-se à elaboração de experiências e pesquisa científica, tais como testes de vacinas, avaliação de terapêuticas e tratamentos, testes de medicamentos etc. Denomina-se este ambiente de ***ambiente micro***, e tem a ver, naturalmente, com a pesquisa laboratorial e científica. Relaciona-se, principalmente por esse motivo, com as disciplinas de ***Imunologia, Fisiologia*** e ***Farmacologia***, dentro do ciclo de formação básica do médico, e com todas as demais áreas

clínicas, em maior ou menor medida, como ***Pediatria, Cardiologia, Neurologia, Pneumologia, Psiquiatria, Gastroenterologia*** etc., toda vez que é indispensável à compreensão da grande maioria das publicações de artigos científicos nessas especialidades.

A Bioestatística, pela sua importância para a pesquisa médica, é disciplina obrigatória da maioria das especialidades de pós-graduação em Medicina.

1.3 População e amostra

Normalmente entende-se o termo população como um conjunto de pessoas. Em estatística, o sentido da palavra se torna mais amplo.

Entende-se por ***população*** a totalidade dos elementos ou de um atributo dos elementos referentes a um conjunto determinado. Assim, é lícito, na linguagem estatística, falar de:

- **População de Itajubá**, cujos elementos são as pessoas que residem nessa cidade.
- **População de pacientes internados no HE-FMIt**, que tem como elementos as pessoas internadas no HE-FMIt (Hospital Escola da Faculdade de Medicina de Itajubá).
- **População de pacientes atendidos no PS do HE-FMIt em 1999**, cujos elementos são as pessoas atendidas no Pronto-Socorro do Hospital..., no ano de 1999.
- **População dos ratos Wistar machos do Biotério da FMIt**, que tem como elementos ratos Wistar machos do...
- **População de seringas descartáveis do Posto de Saúde do Bairro do Cruzeiro**, cujos elementos são seringas...

Em primeiro lugar, nota-se que a população estatística não é necessariamente um conjunto de pessoas. Por outro lado, percebe-se que a população pode ser ***enumerável*** e ***finita***, como o caso do primeiro, terceiro, quarto e quinto exemplos citados. A população pode ser também ***infinita***, ou impossível (difícil) de enumerar, como o caso do segundo exemplo, no qual, por falta de maiores informações a respeito do conjunto, torna-se difícil a sua contagem. Outro exemplo neste sentido seria o das "pessoas portadoras do vírus HIV" Mesmo que esse conjunto fosse (como na verdade é) finito, a sua determinação é de extrema dificuldade, pois, para consegui-la, seria necessário testar todos os habitantes do planeta. Aliás, mesmo que as pessoas fossem de uma cidade, bairro ou até de um grupo, seriam encontradas dificuldades.

A dificuldade em enumerar ou tratar conjuntos completos de dados faz com que se trabalhe com partes do conjunto original, tidas como ***representantes*** do conjunto. Convenciona-se denominar essas partes ***amostras***. Deste modo, uma amostra é uma parte tomada da população, ou um conjunto de elementos da população selecionado segundo algum critério, como será mostrado no item 2.5.1 deste capítulo.

Entendendo a população como um ***conjunto universal***, U, uma amostra pode ser considerada como um subconjunto de U.[1] É claro também que várias amostras podem ser extraídas de uma mesma população, como sugere a Fig. 1.2.

A definição correta da população alvo de um estudo é uma questão de grande importância para a pesquisa médica. Como foi possível observar nos exemplos anteriores, informações acerca da abrangência espacial (geográfica) e temporal servem para definir a população com maior precisão. A rigor, quando é acrescentada a uma coleção de dados uma característica adicional, está-se definindo uma nova população, mais específica (ou menos geral) que a anterior. É importante notar que essa nova população é *outra* população. Por exemplo, imagine que estão sendo estudados os efeitos colaterais de determinado anticoncepcional oral. O objeto de estudo pode ser definido, da forma mais abrangente possível, como a população feminina. Assim, a população seriam todas as mulheres. Contudo, não é sempre fácil ou possível relacionar os efeitos da droga com toda a

---

[1]Como será explicado adiante, se um subconjunto de U contém todos os elementos que comportem um certo atributo, esse subconjunto passa a constituir outra população.

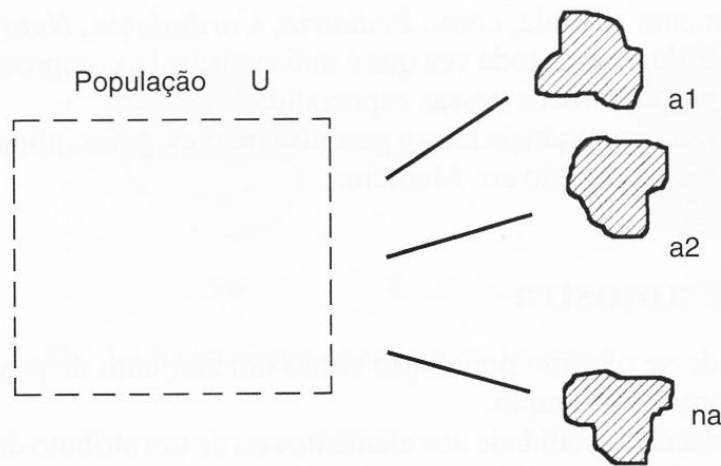



Fig. 1.2 Extração de amostras de uma população.

população. Pode ser, por exemplo, que determinado subgrupo feminino seja mais sensível a determinado efeito. Suponha que esse subgrupo seja o das mulheres hipertensas. Observa-se que, ao se acrescentar uma característica ao grupo "mulheres", tornando-o o grupo "mulheres hipertensas", foi definida uma outra população, diferente da anterior, mesmo que os indivíduos do grupo original estejam necessariamente todos no novo grupo. Não se deve pensar que o novo grupo é uma amostra do primeiro, simplesmente porque é um subconjunto dele. A razão disto é que o grupo "mulheres hipertensas" possui *todas* as mulheres que têm pressão arterial elevada. Portanto, é uma população.

Voltando à questão da definição clara da população: se, no exemplo citado, a experiência usasse o conjunto "mulheres", seria testada uma amostra desse conjunto que iria conter uma proporção de mulheres hipertensas e não-hipertensas. O resultado das conclusões a respeito dos efeitos do anticoncepcional sobre as "mulheres" seria então diferente das conclusões que teriam sido obtidas caso a amostra tivesse sido somente de mulheres hipertensas. Observe que essa informação seria fundamental do ponto de vista clínico. Se uma mulher visita seu ginecologista para se aconselhar sobre o tipo de anticoncepcional mais conveniente para ela, o médico necessitará de informações específicas sobre o fármaco antes de recomendar ou não o seu uso. A Fig. 1.3 ilustra estes conceitos.

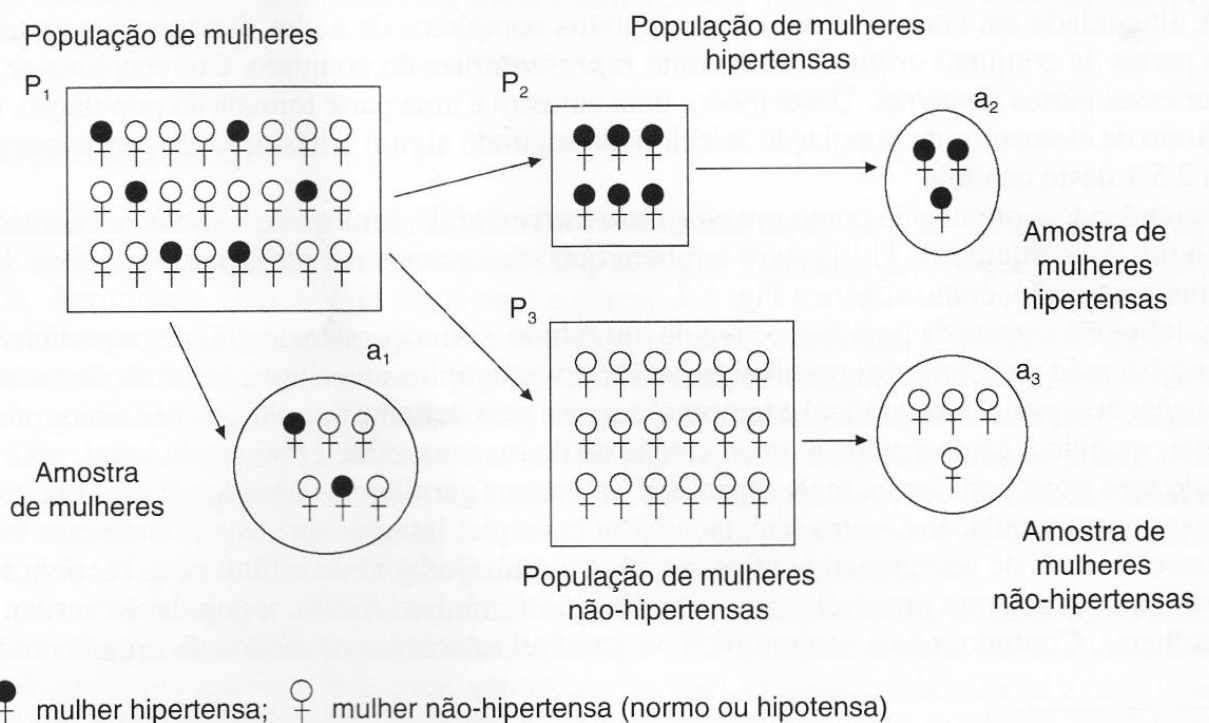



Fig. 1.3 Definição de população e de amostra.

Observe que P_1 representa a população de mulheres, supondo que todas as mulheres estejam incluídas. Ao definir mulheres hipertensas (♀), se forem tomadas todas as mulheres que possuem este atributo, está sendo definida a população representada por P_2. Note todos os símbolos ♀ de P_1 estão em P_2. O mesmo pode ser dito sobre as mulheres não-hipertensas. Já a_2 é uma amostra de P_2, pois contém alguns dos seus elementos. Idêntico raciocínio pode ser aplicado a a_3. Em a_1, vemos que existem elementos de P_1, e que possui proporções casuais ou aleatórias de mulheres hipertensas e não-hipertensas. Note que, eventualmente, uma amostra de P_1 poderia conter indivíduos com apenas um dos atributos, se, por acaso, assim ocorresse.

Tarefa ilustrativa 1.1

Leia a bula de um remédio qualquer e verifique as ***indicações***, as ***contra-indicações*** ou os ***efeitos adversos*** (efeitos colaterais ou efeitos secundários). Observe como as informações estão direcionadas a grupos específicos. Perceba que todas essas informações foram testadas em amostras de cada um dos subgrupos (populações). Verifique se existem restrições de caráter geral, ou seja, indicações sem referência a um grupo específico.

1.4 Dados primários e dados secundários

Ao reunir informações sobre um estudo, o pesquisador normalmente trabalha com dados resultantes de medidas, contagens ou experimentos realizados por ele ou sua equipe. Denominam-se esses dados como ***dados primários***, uma vez que o primeiro a ter acesso aos mesmos é o próprio pesquisador, ou a sua equipe. Considera-se que o pesquisador tenha trabalhado de forma a garantir um nível adequado de qualidade de dados e que as limitações dos mesmos sejam conhecidas. Nos itens 2.3 e 2.4 deste capítulo serão discutidos os problemas subjacentes ao levantamento de dados primários.

Entretanto, são usuais nos estudos os ***dados secundários***, obtidos de diversas fontes, como por exemplo: artigos em periódicos científicos, artigos ou comunicações em eventos científicos ou institutos de pesquisa e estatísticas. Como os dados secundários não foram obtidos diretamente pelo pesquisador e sua equipe, verifica-se que a qualidade deles pode estar prejudicada. Contudo, pelo menos nos exemplos de fontes de dados secundários que foram mencionados, espera-se que a qualidade seja elevada. No caso de periódicos científicos, a qualidade está relacionada normalmente ao prestígio da publicação. Dados ou informações extraídos do *New England Journal of Medicine* ou do *Journal of the American Medical Association*, *JAMA*, por exemplo, são referências mundiais na área de saúde. Dados ou informações extraídas de meios de comunicação leigos, como revistas, jornais ou televisão, costumam ter um nível de qualidade muito baixo. No item 1.6 são mostradas estruturas de dados que podem ser consultadas de fontes de dados secundários para saúde e população confiáveis, referentes a institutos de pesquisa e processamento de dados, como o DataSUS e o Instituto Brasileiro de Geografia e Estatística — IBGE. A Organização Mundial da Saúde — OMS, e a Organização Panamericana de Saúde — OPAS, também costumam auxiliar no levantamento de bons dados secundários para pesquisa.

1.5 Censo

Entende-se Censo como o levantamento ou registro estatístico de uma certa população, de acordo com alguns critérios tais como sexo, idade, religião, estado civil, profissão.

Esta conceituação, entretanto, está relacionada com a definição clássica de Censo, ou com a idéia de Censo Demográfico. Mais modernamente, e de acordo com a definição de população que foi dada anteriormente, a contagem populacional pode estar relacionada ao número de estabelecimentos industriais, rebanhos animais, tamanho de propriedades rurais, número de estabelecimentos bancários etc. Esses censos são denominados: Censo Industrial, Censo Agropecuário e Censo Comercial e de Serviços.

O Censo Demográfico inclui informações relacionadas à população (pessoas) e habitação (residências). Ainda, são investigadas nele o tamanho e composição populacional, estrutura familiar, movimentos migratórios, escolaridade, potencial e qualificação da mão-de-obra, padrões de renda familiar e individual, fecundidade e situação habitacional.

Os censos industrial, comercial, agropecuário e de serviços permitem o levantamento de dados relativos à mão-de-obra empregada, distribuição salarial, produtividade média, capital empregado, estoques, índice de preços etc.

No Brasil, os recenseamentos são efetuados pelo Instituto Brasileiro de Geografia e Estatística — IBGE. O IBGE reúne uma grande quantidade de dados e informações sobre recenseamentos no Brasil, inclusive atuais, que podem ser consultados nas agências do instituto existentes nas principais cidades do País ou no seu *site*: www.ibge.gov.br.

1.6 Estatísticas de saúde no Brasil

Os registros de saúde no Brasil podem ser consultados no *site* do Serviço Único de Saúde (SUS) do Ministério da Saúde: http://www.datasus.gov.br. Esses registros de estatísticas de saúde são elaborados pelo DataSUS e se referem a várias categorias. A seguir é mostrada a estruturação de registros que pode ser ali consultada:

- Assistência à Saúde
- Rede Assistencial
- Morbidade e Informações Epidemiológicas
- Estatísticas Vitais e Mortalidade e Nascidos Vivos
- Recursos Financeiros
- Informações Demográficas e Socioeconômicas

1.6.1 Assistência à saúde

Em termos de assistência à saúde, estão disponíveis séries referentes a:

- Internações por especialidade e local de internação, desde 1981
- Procedimentos hospitalares por local de internação, desde 1992
- Procedimentos hospitalares por local de residência, desde 1995

Produção ambulatorial, imunizações, desde 1994

- Doses aplicadas
- Cobertura

Situação da saúde, desde 1998

- Produção e marcadores
- Cadastramento familiar
- Situação do saneamento

1.6.2 Rede assistencial

Em termos de rede assistencial, estão disponíveis séries referentes a:

- Rede hospitalar, desde 1992
- Rede ambulatorial, desde 1998
- Cadastro Nacional de Estabelecimentos de Saúde
- Pesquisa Assistência Médico-Sanitária (AMS), de 1981 a 1990, 1992, 1999 e 2002.

1.6.3 Morbidade e informações epidemiológicas

Em termos de morbidade e informações epidemiológicas, estão disponíveis séries referentes a:

- Morbidade hospitalar por local de internação, desde 1984
- Morbidade hospitalar por local de residência, desde 1995
- Aids, desde 1980
- Câncer de colo de útero e mama, desde 2002
- Hanseníase, desde 1997
- Saúde bucal, cárie dental, desde 1996

1.6.4 Estatísticas vitais e mortalidade e nascidos vivos

Em termos de Estatísticas vitais e mortalidade e nascidos vivos, estão disponíveis séries referentes a:

- Mortalidade geral, desde 1979
- Nascidos vivos, desde 1994

1.6.5 Recursos financeiros

- Recursos financeiros, AIH e GAP (Guia de Autorização de Pagamento) (de 1990 a 1997)

1.6.6 Informações demográficas e socioeconômicas

Em termos de **Informações demográficas e socioeconômicas** estão disponíveis séries referentes a:

- População
- Educação (alfabetização e escolaridade)
- Abastecimento de água
- Instalações sanitárias
- Coleta de lixo

Embora o DataSUS disponibilize informações demográficas e socioeconômicas, o IBGE disponibiliza uma grande quantidade de informação a este respeito. Os dados mais recentes se referem aos levantamentos do Censo 2000, da Pesquisa Nacional por Amostra de Domicílios de 2002 e da Síntese de Indicadores Sociais, de 2003. Outro tipo importante de informação que pode ser obtida no *site* do IBGE são os dados discriminados por município. Podem ser obtidos dados demográficos e socioeconômicos atualizados para todos os municípios brasileiros.

1.7 Conceitos e informações preliminares importantes

1.7.1 Arredondamento de dados

Muitas vezes o resultado de uma operação contém um grande número de decimais, ou de casas após a vírgula. Geralmente, nesses casos, utiliza-se uma técnica para reduzir o número de algarismos significativos (e o número de casas decimais) para um valor julgado conveniente.

O arredondamento é sempre feito por aproximação. Desta forma, a altura 176,25 cm, arredondada para inteiros (ou sem casas decimais) resulta 176 cm, assim como 176,75 seria 177 cm.

Entretanto, quando o número a arredondar dista (aproxima) igualmente de dois valores possíveis, o critério apresentado não é suficiente. Nesses casos, costuma-se arredondar o número para o par mais próximo, ou para o número par que precede o 5.

Assim, o arredondamento de 176,25 para décimos (ou com uma casa decimal) resultaria 176,2 (pois 0,25 está mais próximo de 0,2 do que de 0,4). Pela mesma razão, 176,75 seria arredondado para 176,8.

Observação: O arredondamento não se refere somente a casas decimais. Pode-se desejar, dependendo do problema, efetuar um arredondamento de valores inteiros. Por exemplo, arredondar o número 1.250.456 para milhares. Neste caso, o arredondamento seria 1.250 milhares.

1.7.2 Classificação de variáveis

Uma ***variável*** pode ser definida como uma característica de um conjunto de elementos, desde que essa característica apresente mais de uma classificação possível. Por esta razão, costuma-se representar uma variável da seguinte forma:

$$X = \{x_1, x_2, \dots, x_j, \dots, x_n\}$$

Onde X representa uma determinada variável (por exemplo, X = peso de camundongos), enquanto $x_1, x_2, \dots, x_j, \dots, x_n$ são observações de X, sendo que pelo menos duas observações são diferentes.

Caso contrário, isto é, se a característica ou o atributo que está sendo aferido admite uma única classificação, ter-se-á uma ***constante***. Por esta razão, as constantes costumam ser representadas matematicamente por uma única letra. Por exemplo, a constante a.

As variáveis, de acordo com a forma como são medidas, admitem a seguinte classificação:

1. Variáveis quantitativas ou numéricas: são avaliadas de forma numérica.
2. Variáveis qualitativas ou literais: são expressas por a) palavras (atributos), como por exemplo, tipo sanguíneo, ou b) ordens, como 1.°, 2.° etc.

Por sua vez, as variáveis quantitativas se dividem em:

1.1. Discretas: admitem somente números inteiros (conjunto dos números Naturais). Exemplos seriam as variáveis resultantes de contagens, como o número de batimentos cardíacos, a freqüência respiratória ou a concentração de plaquetas no sangue.

1.2. Contínuas: admitem números fracionários (conjunto dos números Reais). Estão relacionadas com mensurações. Tempo de coagulação, perímetro cefálico e peso do fígado são exemplos de variáveis contínuas.

Um caso especial das variáveis qualitativas é o das que possuem somente duas categorias — as chamadas variáveis binárias ou dicotômicas. Esse tipo de variável somente assume um de dois possíveis valores — zero ou um. Normalmente, a inclusão de um sintoma como variável em um estudo é feita de forma binária, considerando a existência ou não do mesmo. Por exemplo, a variável "dor de cabeça", admitindo as possibilidades de resposta: "sim" ou "não"

Embora algumas variáveis tenham uma classificação clara, como, por exemplo, o "tipo sanguíneo", muitas outras variáveis admitem ser tratadas de forma qualitativa ou quantitativa, dependendo do caso. Por exemplo, considere a pressão arterial. Em um estudo, essa variável pode ser numérica, empregando medidas expressas em milímetros de mercúrio (mm Hg), ou qualitativa, classificando a pressão como: "baixa", "normal" "ou elevada", ou ainda binária: hipertenso, não-hipertenso.

2 A Preparação de um Estudo

A preparação de um estudo é uma fase fundamental da pesquisa científica. Diversos procedimentos estatísticos estão presentes nessa fase, de forma que é importante que se considere a participação de um bioestatístico durante as discussões preliminares dos projetos de pesquisa, ou que se tenha no grupo um pesquisador experiente que tenha uma visão geral dos principais métodos estatísticos relacionados com a área. Este cuidado costuma evitar ulteriores frustrações em relação aos resultados alcançados pela pesquisa.

2.1 Tipos de estudos

Em princípio, pode-se dizer que os estudos científicos se dividem em dois grandes grupos: i) os ***estudos observacionais*** e ii) os ***estudos experimentais***.

Os estudos observacionais caracterizam-se pela não intervenção do pesquisador no ambiente da pesquisa. Neles, as situações ocorrem naturalmente e o pesquisador observa as características dos pesquisados e faz comparações e descrições. Já nos estudos experimentais, o pesquisador produz uma situação artificial, na qual ele aloca aleatoriamente os elementos pesquisados a grupos (de modo a serem submetidos ou não a uma vacina, um medicamento ou outro procedimento), a fim de terem os efeitos avaliados em condições controladas de observação.

Na literatura médica, os principais desenhos de estudos encontrados são:

- Estudo retrospectivo ou de casos e controles
- Estudo de coortes
- Estudo de prevalência ou transversal
- Estudo ecológico

} Estudos observacionais

- Estudo experimental — ensaio clínico aleatório.

Estudo Retrospectivo ou de Casos e Controles

A característica específica dos estudos retrospectivos é a de que estes são iniciados após os indivíduos terem desenvolvido a condição (patológica ou não) investigada. Esses estudos regridem no tempo para determinar as características que os indivíduos apresentavam antes do início da condição. Nos estudos de casos e controles, os "casos" são os indivíduos que desenvolveram a condição, enquanto os que não a desenvolveram são os "controles".

Estudo de Coortes (grupos)

Coorte é um grupo de indivíduos que compartem uma experiência. Nesses estudos, são formados dois coortes, um que possui as características estudadas e um outro coorte que não as possui.

Estudo de Prevalência ou Transversal

Esse tipo de estudo consiste em estudar determinada população, ou uma amostra representativa dela, em função de apresentar características que possibilitem a investigação exposição-doença. Nesse tipo de estudo, as informações sobre exposições e doenças da população são medidas simultaneamente em um curto período de tempo. Assim, o estudo transversal fornece um retrato de como as variáveis estavam relacionadas no momento da pesquisa.

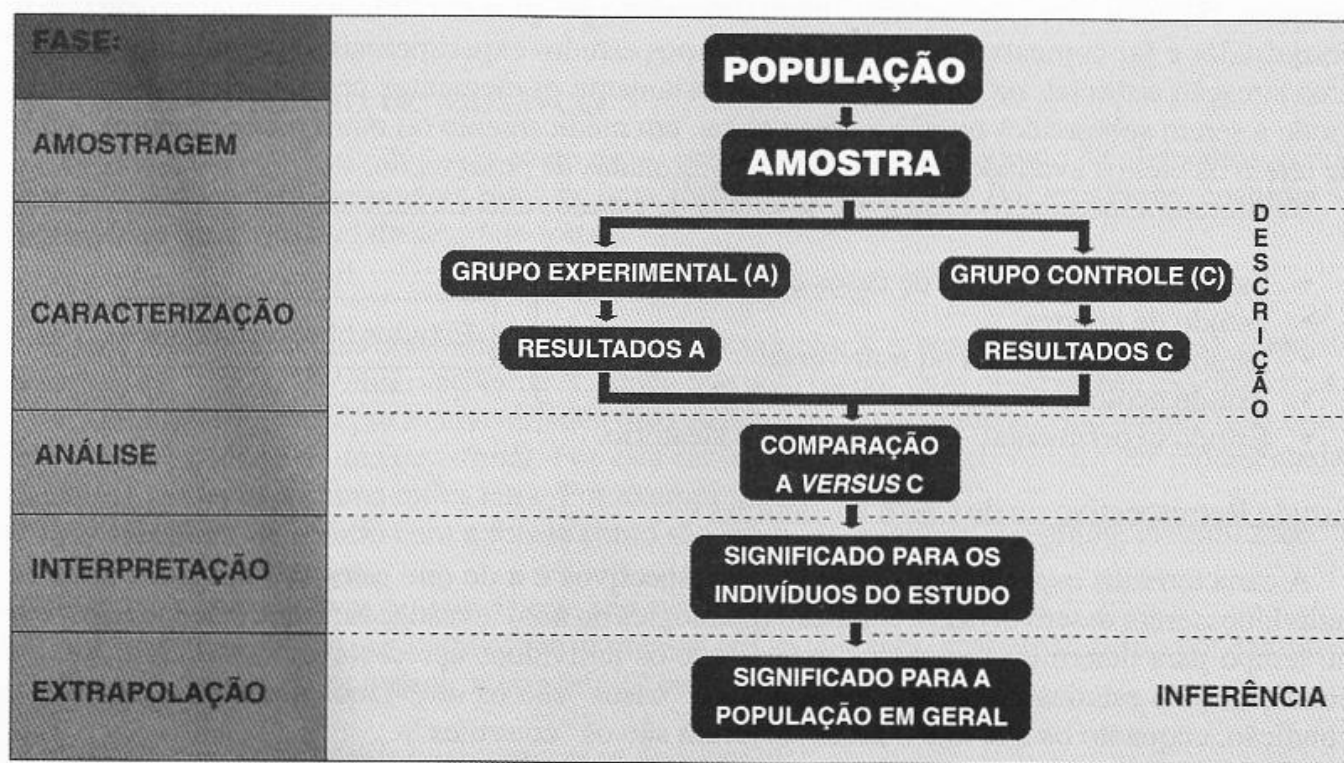
Estudo Ecológico

Nesse delineamento a unidade de análise não é constituída de indivíduos, mas de grupos de indivíduos. Assim, não se sabe se um determinado indivíduo da população investigada é exposto ou doente; apenas as informações globais estão disponíveis, como a proporção ou freqüência da doença na população e a proporção ou freqüência da exposição na população.

Ensaio Clínico Aleatório

Esse ensaio, também chamado de controlado, é um experimento no qual os indivíduos ou elementos são alocados aleatoriamente para grupos, de estudo (ou também chamado experimental) e controle (testemunha) de modo a serem submetidos ou não a uma vacina, um procedimento, um medicamento, e terem seus efeitos avaliados em condições controladas de observação. Em

condições ideais, os indivíduos são escolhidos ao acaso e às cegas, tanto no grupo controle quanto no grupo tratamento.[2]

Esquema Geral de um Estudo Clínico

2.2 Escolha das variáveis ou dos fatores

Existem vários critérios para auxiliar na hora de decidir quais as variáveis ou fatores que deverão ser incluídos no estudo. Porém, essa escolha está intimamente ligada ao tipo de estudo a ser realizado, ao assunto de interesse e à sua respectiva literatura. A pesquisa bibliográfica nas bases de dados existentes é imprescindível para auxiliar no levantamento das variáveis que já foram relacionadas ao assunto de interesse da pesquisa. Essas variáveis devem ser incorporadas no ambiente de pesquisa atual para controle de possíveis vieses. De uma maneira geral, as variáveis que devem ser consideradas em uma pesquisa podem ser agrupadas em:

1. Fatores de interesse direto (diretamente relacionados *a priori*).
2. Fatores que podem modificar a ação dos fatores principais (a), ou que podem elucidar o funcionamento dos fatores principais (b).
3. Fatores relacionados com a técnica empregada no experimento.
4. Fatores de classificação, sugeridos por agrupamentos naturais das unidades experimentais.
5. Variações propositais introduzidas nas unidades experimentais.

Considere, por exemplo, um estudo sobre doenças respiratórias. Alguns exemplos de fatores do primeiro grupo poderiam ser: i) presença de pigarro, ii) tosse freqüente, iii) dores de garganta. Estas variáveis servem para avaliar o estado do paciente, em função da presença ou não do

[2]Amostragem aleatória significa que qualquer indivíduo tem uma probabilidade conhecida de ser selecionado para qualquer um dos grupos. Será discutida no item 2.5.1 deste capítulo. Escolha "às cegas" ocorre quando nem pacientes nem pesquisadores saibam em qual dos grupos se encontra um determinado participante do experimento. Esse método será discutido no item 2.4.

sintoma. Observe que as variáveis exemplificadas são de natureza qualitativa, porém, variáveis quantitativas poderiam ser incluídas, dependendo do caso.

Em relação a fatores da categoria 2.a, podem ser selecionadas variáveis como: i) tabagismo, ou ii) qualidade do ar. No caso 2.b, idade.

Em relação aos fatores da categoria 3, teria que ser definido o tipo de experimento. Se, por exemplo, os dados estão sendo obtidos efetuando exames de pacientes por dois ou três médicos, a informação sobre qual deles realizou o exame em cada paciente poderia ser importante. O fator seria, por exemplo "exame", e 1, 2 e 3 seriam os médicos que realizaram as consultas.

Fatores de classificação, sugeridos por agrupamentos naturais das unidades experimentais seriam, por exemplo: i) sexo, ii) faixa etária, iii) etnia etc.

Variações introduzidas no estudo poderiam ser: i) a introdução de um grupo de indivíduos com asma brônquica ou ii) tratar uma parte do grupo com um descongestionante. O objetivo seria o de verificar as alterações provocadas nos indivíduos ou no experimento em geral.

2.3 Problemas usuais no levantamento de dados

Um conceito importante que se deve ter em mente quando se trata de levantamento de dados é o de que o sucesso nas conclusões tiradas acerca da população com base nas informações colhidas de uma ou mais amostras depende da criteriosa seleção desta(s). Assim, dados mal coletados irão carregar suas distorções para qualquer análise que se faça deles.

Toda pesquisa tem limitações, nos seus resultados, decorrentes da metodologia empregada na investigação. O sucesso nas conclusões acerca da população, com base nas informações colhidas de uma ou mais amostras, depende do grau com que o estudo atinge conclusão correta, ou seja, da validade do estudo. Questões metodológicas não resolvidas no decorrer da pesquisa, no modo de seleção das pessoas, na forma de obtenção dos dados ou na maneira de analisá-los constituem ameaças à validade, à credibilidade das conclusões.

Basicamente existem dois tipos de validade: validade interna e validade externa.

A validade interna diz se as conclusões de uma pesquisa são corretas para a amostra estudada. Essa validade é pré-requisito para a validade externa.

A validade externa diz se as conclusões de uma pesquisa são aplicáveis à população da qual a amostra se originou ou às outras populações. Essa validade possui dois aspectos distintos de extrapolação de resultados:

- Extrapolação da amostra para a população da qual a amostra teve origem. O sucesso da generalização depende da observância das regras estatísticas de amostragem e está relacionado à representatividade dos participantes incluídos na amostra, em face da população da qual foi retirada esta amostra.
- Extrapolação da população investigada para outras populações. Os resultados de uma investigação, já generalizados para uma determinada população, são utilizados para outras populações que não disponham de informações semelhantes.

A limitação dos resultados de uma pesquisa científica deve ser discutida à luz dos possíveis erros metodológicos (vieses) que constituem ameaças à validade da pesquisa. Busca-se, então, um nível de confiança que é possível depositar nos resultados e conclusões a partir de uma análise criteriosa dos possíveis erros e suas causas.

Denomina-se viés um erro sistemático (vício, distorção) e não intencional, proveniente de questões metodológicas.

Existem dois fatores que influenciam no correto levantamento dos dados: a representatividade e a fidedignidade.

2.3.1 Representatividade

A representatividade é um fator associado à forma de amostragem. Assim, de modo geral, quando se trata de selecionar uma amostra, procuram-se reproduzir as características observáveis da população.[3] Este procedimento é conhecido como ***critério de proporcionalidade*** e, quando ele é considerado, diz-se que a amostra é ***representativa*** da população em questão.

Entretanto, para que isto seja possível, as características da população devem ser previamente conhecidas. Contudo, a disponibilidade de informações prévias sobre a estrutura populacional para um estudo nem sempre é a ideal, de forma que, normalmente, a fonte de informação acaba sendo um outro estudo ou dados estatísticos secundários, extraídos de institutos de pesquisa, por exemplo.

Na falta de informações sobre a composição da população em estudo, trata-se de que a escolha seja a mais isenta possível, adotando-se algum critério de ***aleatoriedade*** (escolha ao acaso), como um sorteio.

Todavia, quando os dados necessários para o conhecimento dos atributos da população estão disponíveis e são ignorados ou manipulados, a amostra resultante seria considerada ***tendenciosa***. Conclusões e estimativas efetuadas com base nessa amostra não possuiriam ***consistência***. (Ver Fig. 1.4.)

EXEMPLO 1.1

Tomando um exemplo para ilustrar a questão, imagine um estudo sobre a taxa de *coliformes* presentes na água consumida num determinado município. Parece claro que o estudo deverá ser feito por amostragem domiciliar da água, pois seria inviável testar todos os domicílios. Suponha ainda que existem outras informações como: i) a classificação das áreas geográficas (rural e urbana, por exemplo), ii) a localização dos bairros, iii) a população dos bairros e iv) a taxa de atendimento com água tratada. Veja o Quadro 1.1.

Quadro 1.1 Exemplo de amostragem com informações

População	Número de Habitantes	Taxa de Atendimento com Água Tratada
Rural	**1.000 hab.**	**0%**
Urbana	**9.000 hab.**	**89%**
Bairro A	2.000 hab.	100%
Bairro B	2.000 hab.	100%
Bairro C	5.000 hab.	80%

Contando com essas informações, uma amostra bem selecionada e, portanto, ***representativa*** da população de domicílios deveria observar que:

- todos os locais sejam alcançados pela pesquisa, observando-se a distribuição geográfica (área urbana — com seus três bairros — e área rural). Não é possível, por simples conveniência, efetuar a pesquisa somente em domicílios urbanos ou, pior, no bairro A;
- seja considerada, nesse caso, a população de cada setor geográfico. A amostra deverá ser constituída de partes proporcionais idênticas às da população. Assim, 10% da amostra seriam constituídos por domicílios rurais, 20% do bairro A etc. Na falta das populações, poderiam ser empregados pesos para a constituição da amostra, de acordo com o sentimento do analista;
- exista total aleatoriedade na escolha dos domicílios.

Caso contrário, poder-se-ia dizer que a amostra é ***tendenciosa*** e que, portanto, as conclusões dela decorrentes não possuem ***consistência***.

[3]Por exemplo, se é sabido que, na população em estudo, 20% dos indivíduos são hipertensos, seria conveniente que uma amostra dessa população contivesse também 20% de indivíduos com esse atributo.

2.3.2 Fidedignidade

Outro aspecto que deve ser levado em conta no trabalho de levantamento de dados e constituição das amostras é o da ***fidedignidade***. A fidedignidade dos dados ou das informações está relacionada com a ***precisão*** dos dados, ou a sua qualidade. Veja o Quadro 1.2.

A falta de precisão pode ser ocasionada por vários motivos. De modo geral, podem ser citados: falhas nos instrumentos de aferição dos dados, problemas nos questionários usados para o levantamento de dados e falhas humanas.

No primeiro caso, o uso de balanças ou esfigmomanômetros mal calibrados, réguas ou estetoscópios defeituosos ou em mau estado de conservação pode provocar erros nas medidas. Também alguns exames laboratoriais, como os que implicam contagem, estão sujeitos à falha. Nesses casos, durante o levantamento dos dados da amostra, recomenda-se usar sempre os mesmos aparelhos.

No segundo caso, existem informações que carregam grande margem de erro. Por exemplo, suponha que um questionário contenha a seguinte pergunta: Quanto tempo faz desde que você foi ao médico pela última vez?

Mesmo que o entrevistado tenha boa vontade em responder e não esteja tencionando falsear a resposta, ele pode encontrar dificuldades em realmente precisar o tempo solicitado pela pergunta. A razão da resposta do entrevistado poder apresentar uma falha de informação é simplesmente o fato de que um conjunto de informações não é devidamente registrado e, para obtê-las, às vezes, não existe outra alternativa a não ser contar com a memória do entrevistado.

Outro exemplo de uma questão que poderia suscitar erros seria: Qual é a freqüência do seu relacionamento sexual?

Tratando-se de um tema que em boa parte dos casos cria constrangimento ao entrevistado, ele pode falsear a sua resposta, exagerando ou escondendo a verdadeira informação.

Não é difícil imaginar outras questões incômodas, tais como: Qual a sua renda? Faz higiene bucal diariamente? Já dirigiu embriagado?

Pelos mais variados motivos, desde a precária memória em alguns casos, passando pelo medo e a própria fantasia do entrevistado, é comum a falta de precisão das respostas. Por este motivo, os questionários de levantamento de dados devem ser elaborados com extremo cuidado, evitando perguntas vagas ou que dêem margem a respostas muito subjetivas. Naturalmente, seus resultados devem ser tomados com grande cautela.

Quadro 1.2 Resultados da seleção da amostra quanto à fidedignidade

Fidedignidade	Fator Associado	Estado	Precisão
QUALIDADE DOS DADOS	Instrumentos de aferição	Corretamente calibrados	↑
		Falhas nos instrumentos	↓
	Processo de aferição	Competência técnica	↑
		Falhas humanas	↓
	Questionários	Bem elaborados	↑
		Falhas nos questionários	↓

2.4 Efeitos indesejados no levantamento de dados

2.4.1 Efeito placebo

O efeito placebo ocorre quando um indivíduo participante de um experimento, mas não tratado, acredita estar recebendo o tratamento e passa a relatar melhoras em seus sintomas.

Placebo é uma substância neutra, isto é, que não apresenta nenhum princípio ativo capaz de provocar alterações no estado de um indivíduo. Uma forma de administrar placebo por via oral em comprimidos consiste em dar ao indivíduo um comprimido de farinha. Outra forma é apresentar a substância diluída em uma bebida, como suco ou chá. Normalmente, os comprimidos de placebo possuem o mesmo formato dos comprimidos verdadeiros, isto é, aqueles que contêm o princípio ativo. A intenção com esse procedimento é a de que o paciente, ao tomar o placebo, de fato pense estar tomando uma substância que irá lhe trazer uma mudança de estado. Naturalmente, nada deveria ocorrer com ele e, se relatar melhora, obviamente está sendo objeto do efeito placebo.[4] Esse procedimento é de fundamental importância para isolar a ação do sal ao se testar a sua eficácia. O procedimento experimental empregado para testar o efeito de determinada substância eliminando o efeito placebo consiste em separar os pacientes em dois grupos. O primeiro — denominado controle — recebe comprimidos de placebo; enquanto o segundo — denominado tratamento — recebe os comprimidos verdadeiros. Devido ao fato de os pacientes participantes do experimento não saberem ao certo se estão sendo tratados (recebendo o comprimido verdadeiro), um estudo que adota esse procedimento é denominado ***estudo cego***, ou *blind*.

2.4.2 Efeito Rosenthall

Denomina-se efeito Rosenthall ou efeito do experimentador a qualquer mudança ou alteração do padrão de resposta do indivíduo pesquisado provocada, involuntariamente, pelo pesquisador ou experimentador. A influência pode manifestar-se pelo comportamento do experimentador, através do tom da sua voz, sua expressão facial ou a própria atitude.

Esse tipo de efeito é muito comum em pesquisas com questionários. Um exemplo seria perguntar ao entrevistado, com uma entonação de censura, sobre determinada informação. A inibição que pode resultar no entrevistado poderia mudar a sua maneira de responder.

O efeito Rosenthall também pode manifestar-se em relação ao teste de substâncias ativas. Isto pode ocorrer por ocasião da administração da substância se, de alguma forma, o paciente desconfia, pela atitude do experimentador, estar recebendo ou não o sal. Para evitar a ocorrência do efeito Rosenthall nos testes com sais, é necessário evitar que o experimentador saiba qual tipo de substância (ativa ou placebo) está sendo administrada ao paciente. Assim, de forma análoga ao caso do efeito placebo, deve-se cegar o experimentador. O experimento onde nem o experimentador nem o paciente sabem distinguir o comprimido verdadeiro do comprimido de placebo é denominado ***estudo duplo cego***, ou *double blind*. Veja o Quadro 1.3. Devido à capacidade de neutralizar os efeitos placebo e Rosenthall, os estudos do tipo duplo cego fazem parte do protocolo de teste de substâncias na indústria farmacêutica e nos organismos de controle de drogas.

Quadro 1.3 Estudos e efeitos tratados

Experimento	Paciente	Pesquisador	Estudo	Efeito Tratado
CONHECIMENTO	NÃO	SIM	Cego (*blind*)	Placebo
	NÃO	NÃO	Duplo cego (*double blind*)	Placebo + Experimentador

[4]Apesar de que, ao tomar placebo, o indivíduo realmente não tomou substância alguma capaz de mudar seu estado, isto não quer dizer que, ao relatar uma melhora, o indivíduo esteja mentindo. As melhoras realmente podem ocorrer. As razões para esse fato não são bem compreendidas, embora se possa especular sobre a ativação inconsciente de mecanismos internos de alívio, defesa ou cura do organismo.

2.4.3 Efeito Hawthorne

O efeito Hawthorne ocorre quando um indivíduo participante de um experimento e tratado passa a responder de forma diferente pelo fato de estar participando da experiência. Uma explicação para esse comportamento seria o estado de ansiedade provocado pela nova situação vivida pelo experimentado. Assim, a simples mudança de disposição ou de ânimo do paciente poderiam constituir potenciais fontes de alteração dos resultados obtidos no experimento ou na entrevista, caso se trate de uma pesquisa por questionário. O efeito Hawthorne também pode ser diminuído em um experimento empregando estudos do tipo duplo cego, porque os pacientes não sabem se realmente estão tomando os comprimidos que contêm o princípio ativo.

2.4.4 Ilusão de Müller-Lyer

A ilusão de Müller-Lyer se verifica quando a simples alteração da ordem das questões de uma pesquisa altera as respostas dos pesquisados. Tal comportamento pode ser decorrente da tendência do entrevistado em ir se adaptando à situação à medida que vai respondendo um questionário. De certa forma, o entrevistado vai tentando formar uma idéia dos propósitos da pesquisa à qual está sendo submetido, talvez com a intenção de evitar situações que o coloquem em uma situação incômoda. Por esta razão, ao alterar a ordem das questões, está-se alterando a percepção do entrevistado e, possivelmente, sua reação ou resposta.

EXEMPLO 1.2

A ordem das questões pode alterar o resultado das entrevistas (ilusão de Müller-Lyer). Suponha um questionário com as seguintes perguntas:

1. Considera a ingestão excessiva de açúcar prejudicial à sua saúde?
2. Consome muito açúcar?

Existem aqui dois aspectos a considerar. Quando as questões são efetuadas na ordem 1-2, o entrevistado é "obrigado" a: i) estabelecer uma relação entre as variáveis *ingestão de açúcar* e *saúde* e ii) efetuar um julgamento sobre a ingestão de açúcar (questão 1) antes de declarar se consome ou não muito açúcar (questão 2). Quando a ordem das questões é 2-1, ele responde a primeira questão de forma espontânea. Desta forma, parece que a ordem 2-1 leva a respostas mais "honestas".

2.4.5 Efeito complacência

O efeito complacência está relacionado com a tendência que os entrevistados apresentam de responder de forma positiva às questões. Assim, eles tendem a manter uma concordância com a questão à qual estão sendo submetidos, divergindo dela em menor medida.

EXEMPLO 1.3

Como exemplo do efeito complacência, imagine a seguinte situação: Um grupo de pessoas é perguntado a respeito do comportamento recente de um determinado dirigente sindical.

Considere a seguinte pergunta:

1. Concorda que o dirigente agiu de forma honesta?

Se a pergunta 1 fosse feita de forma diferente, porém equivalente:

2. Concorda que o dirigente agiu de forma desonesta?

Devido ao efeito complacência, o percentual de indivíduos que considera a atuação do dirigente como honesta (ou seu equivalente, que seria "não desonesta") deveria ser diferente. Se o efeito concordância não existisse, honesto e não desonesto deveriam ser equivalentes.

2.4.6 Efeito memória

O efeito memória ocorre quando a ordem dos itens das questões pode alterar os resultados de um levantamento de dados. Este efeito está provavelmente associado à variação da capacidade de atenção e compreensão do receptor durante o período em que está sendo elaborada a questão. Assim, inicialmente, o entrevistado está com toda a atenção voltada à questão, mas, à medida que a mesma vai sendo elaborada, sua atenção diminui devido à necessidade de processar a informação contida na pergunta. Naturalmente, quanto mais longa a pergunta, maior a exposição a este tipo de efeito.

EXEMPLO 1.4

Suponha que, em um estudo sobre o nível de informação do público, se deseje saber a opinião das pessoas quanto aos principais fatores de prejuízo à saúde. Considere as seguintes perguntas:

1. Considera que o **tabagismo** contribui mais do que o **alcoolismo** para a piora das condições de saúde da população?
2. Considera que o **alcoolismo** contribui mais do que o **tabagismo** para a piora das condições de saúde da população?

Embora as perguntas sejam equivalentes e, conseqüentemente, devessem levar aos mesmos resultados quanto à opinião das pessoas, quando a pergunta 1 é efetuada, existe uma tendência de se considerar mais o primeiro fator perguntado, isto é, o tabagismo. Desta forma, quando a pergunta feita é a 2, o fator alcoolismo passa a ser mais lembrado.

2.4.7 Efeito desmascaramento

O efeito desmascaramento ocorre pela desconfiança do entrevistado quanto à manutenção do sigilo das suas informações e da sua identidade. Evidentemente, o efeito desmascaramento se torna mais forte quando o entrevistado é solicitado a responder questões que ele considera que, de alguma maneira, comprometem a sua imagem. Assim, por exemplo, um indivíduo que preserva uma imagem conservadora poderia sentir-se constrangido em responder a perguntas sobre temas considerados polêmicos, dando respostas contrárias ao que supostamente ele deveria opinar. Embora não se trate de pesquisas por questionários, os efeitos do desmascaramento podem ser bem entendidos lembrando de votações ou eleições. O princípio do voto secreto tem justamente a finalidade de preservar os votantes, de forma a permitir que expressem livremente sua opinião.

2.5 Obtenção de amostras

Basicamente, existem dois problemas relacionados com a obtenção de amostras:

- as formas de realizar a amostragem;
- a determinação do tamanho da amostra.